Journal of Mathematical Inequalities
Volume 6, Number 3 (2012), 501–507 doi:10.7153/jmi-06-47

AN INEQUALITY FOR THE BINARY ENTROPY FUNCTION AND AN APPLICATION TO BINOMIAL COEFFICIENTS

VANIA MASCIONI

(*Communicated by A. Horwitz*)

Abstract. We prove a tight inequality of non-exponential type for a weighted geometric mean commonly appearing when using Stirling's approximation (also frequently studied in its logarithmic form when computing entropies). As an application we prove corollaries involving binomial coefficients.

1. The main inequalities

If $r \in \mathbb{R} \setminus \{0\}$, in the notation from Mitrinović's classic text [1, 2.14], the general *weighted mean of order* r of a sequence of positive numbers $a = (a_1, \dots, a_n)$ with positive weights $w = (w_1, \dots, w_n)$ is defined as

$$M^{[r]}(a,w) := \left(\frac{\sum_{i=1}^n w_i a_i^r}{\sum_{i=1}^n w_i} \right)^{1/r}.$$

The *weighted geometric* mean (corresponding to the limit $r \to 0$ in the above definition) is defined as

$$M^{[0]}(a,w) := \left(\prod_{i=1}^n a_i^{w_i} \right)^{1/\sum_{i=1}^n w_i}.$$

Further, as customary we can set $M^{[-\infty]}(a,w) := \min_i a_i$ and $M^{[\infty]}(a,w) := \max_i a_i$. $M^{[-1]}(a,w)$ is referred to as the *weighted harmonic mean*, while $M^{[1]}(a,w)$ is the *weighted arithmetic mean*. It is well-known [1, 2.14.2,Th. 1] that unless all a_i are identical the function $r \to M^{[r]}(a,w)$ is strictly increasing.

From the point of view of weighted means, the binary entropy function (for the probability p of a coin toss, see for example [2]) is the negative logarithm in base 2 of $p^p(1-p)^{1-p}$, that is, in the above notation, of $M^{[0]}((p,1-p),(p,1-p))$. Our main result is a very tight inequality for this weighted geometric mean (first inequality in Theorems 1 and 2) whose interest lies in its simplicity (not containing exponentials) and thus possibly lending itself to streamlining further calculations.

We provide the second inequality in Theorem 1 (which is actually harder to prove) as it may be of independent interest but, mostly, because it underlines the tightness of the first one given that the right hand side term in Theorem 1 is a weighted average from the same family as the left hand side.

Mathematics subject classification (2010): Primary: 26E60, Secondary: 26D15,05A10, 26D20.
Keywords and phrases: Weighted geometric mean, binary entropy function, Stirling approximation, binomial coefficient.

Paper JMI-06-47

THEOREM 1. *For $q \leqslant 0$, $p \geqslant \frac{1}{3}$ and $a,b > 0$ we have*

$$M^{[q]}((a,b),(a,b)) \leqslant \frac{(a+b)^3}{(a+b)^2+4ab} \leqslant M^{[p]}((a,b),(a,b)) \tag{1}$$

with identity if and only if $a = b$ (or if one takes limits and allows for a and/or b to be zero). Further, the values $q = 0$ and $p = \frac{1}{3}$ are optimal.

These inequalities (for $q = 0$ and $p = \frac{1}{3}$) are surprisingly tight, so much that a quick look at the three simultaneous 3D-plots in *Mathematica* may give the illusion of a single sheet. Since it's enough to consider the case where $a+b = 1$, the monotonic behavior of the means shows that we only need to prove the following more explicit statement:

THEOREM 2. *For $x \in [0,1]$ (we define $x^x|_{x=0}$ and $(1-x)^{1-x}|_{x=1}$ to be 1), we have*

$$x^x(1-x)^{1-x} \leqslant \frac{1}{1+4x-4x^2} \leqslant \left(x^{4/3}+(1-x)^{4/3}\right)^3 \tag{2}$$

where identities hold if and only if $x \in \{0, \frac{1}{2}, 1\}$.

We delay the proofs to the appendix (the proof of the optimality of $q = 0$ and $p = \frac{1}{3}$ in (1) is to be found at the very end).

2. An application to estimates of binomial sums

It is well-known that factorials (and thus binomial coefficients) allow for exact expressions in terms of trigonometric integrals. For example, we have the classical result

$$\binom{n}{k} = \frac{2^{n-1}}{\pi}\int_{-\pi}^{\pi}\cos((2k-n)t)\cos^n t\,dt$$

which can be obtained by simply converting the cosines in the integrand into complex exponentials via Euler's formula, and then expanding using the binomial theorem. In contrast with this kind of result, as an application of the first inequality in Theorem 2 (and indeed the application was the first motivation behind the work on the inequality) we prove estimates for binomial coefficients (see the next Lemma and Theorem 4), where the interest lies in a particularly simple connection of the estimate with the ratio k/n of the integers appearing in $\binom{n}{k}$. Before proving the estimate for sums, we isolate a core idea to show how Theorem 2 gets used. For an alternate and recent approach to estimating binomial coefficients we refer to the excellent article [4].

LEMMA 3. *Let $0 < k < n$. Writing $\alpha := k/n$ we have:*

1. $\dfrac{\binom{n}{k}}{2^n\sqrt{2\pi\alpha(1-\alpha)}} > \exp\left(\frac{1}{1+12n} - \frac{1}{12\alpha(1-\alpha)n}\right)\left(\frac{1}{2}+2\alpha(1-\alpha)\right)^n$
2. $\dfrac{\binom{n}{k}}{2^n\sqrt{2\pi n\alpha(1-\alpha)}} < \exp\left(\frac{1}{12n} - \frac{2(1+6n)}{1+12n+144\alpha(1-\alpha)n^2}\right)\left(\frac{1}{2}+\sqrt{\alpha(1-\alpha)}\right)^n.$

Proof. Robbins [3] proved the following refinement of the classical Stirling's approximation to the factorial:

$$m! = \sqrt{2\pi m}\cdot m^m e^{-m}\cdot e^{r_m}$$

where

$$\frac{1}{1+12m} < r_m < \frac{1}{12m}$$

for $m \geqslant 1$. For integers $0<k<n$ set $k=\alpha n$ (thus $\alpha\in(0,1)$) and apply Robbins' inequalities to obtain

$$\exp\left(\frac{1}{1+12n}-\frac{1}{12\alpha(1-\alpha)n}\right) < \binom{n}{k}\sqrt{2\pi n\alpha(1-\alpha)}\left(\alpha^{\alpha}(1-\alpha)^{1-\alpha}\right)^n$$
$$< \exp\left(\frac{1}{12n}-\frac{1}{1+12\alpha n}-\frac{1}{1+12(1-\alpha)n}\right).$$

We can rewrite the first inequality in (2) as

$$\frac{1}{\left(\alpha^{\alpha}(1-\alpha)^{1-\alpha}\right)^n} \geqslant 2^n\left(1-2\left(\alpha-\frac{1}{2}\right)^2\right)^n,$$

and thus we have a lower estimate

$$\frac{\binom{n}{k}}{2^n\sqrt{2\pi n\alpha(1-\alpha)}} > \exp\left(\frac{1}{1+12n}-\frac{1}{12\alpha(1-\alpha)n}\right)\left(1-2\left(\alpha-\frac{1}{2}\right)^2\right)^n.$$

To obtain the upper estimate we can make use of the inequality

$$M^{[-1/2]}\left((\alpha,1-\alpha),(\alpha,1-\alpha)\right) = \frac{1}{1+2\sqrt{\alpha(1-\alpha)}} \leqslant \alpha^{\alpha}(1-\alpha)^{1-\alpha}$$

(which trivially follows from the monotonic behavior of weighted means [1]). Thus, arguing as above from (3) we have

$$\frac{\binom{n}{k}}{2^n\sqrt{2\pi n\alpha(1-\alpha)}} < \exp\left(\frac{1}{12n}-\frac{1}{1+12\alpha n}-\frac{1}{1+12(1-\alpha)n}\right)\left(\frac{1}{2}+\sqrt{\alpha(1-\alpha)}\right)^n,$$

as stated in the lemma. □

THEOREM 4. *We have the following estimates:*

1. *For integers $2\leqslant k\leqslant \frac{n}{2}$:*

$$\sum_{k=k_1}^{k_2}\binom{n}{k} > 2^n\exp\left(\frac{1}{1+12n}-\frac{n}{12(n-1)}\right)\sqrt{\frac{n}{\pi}}\int_{\beta_1}^{\beta_2}\cos^{2n}t\,dt,$$

where $\beta_1=\arcsin\left(\sqrt{2}\left(\frac{k_1-1}{n}-\frac{1}{2}\right)\right)$ and $\beta_2=\arcsin\left(\sqrt{2}\left(\frac{k_2}{n}-\frac{1}{2}\right)\right)$, and $n\geqslant 1$ (note that $\beta_j\in(-\pi/4,0)$).

2. *For integers* $1 \leqslant k_1 \leqslant k_2 < \frac{n}{2}$*:*

$$\sum_{k=k_1}^{k_2} \binom{n}{k} < 2^n \exp\left(\frac{1}{12n} - \frac{2}{1+6n}\right) \sqrt{\frac{2n}{\pi}} \int_{\gamma_1}^{\gamma_2} \cos\left(\frac{t}{2}\right)^{2n} dt$$

where $\gamma_1 = \arcsin\left(2\left(\frac{k_1}{n} - \frac{1}{2}\right)\right)$ *and* $\gamma_2 = \arcsin\left(2\left(\frac{k_2+1}{n} - \frac{1}{2}\right)\right)$*, and* $n \geqslant 1$ *(note that* $\gamma_j \in (-\pi/2, 0)$*).*

As it will be clear from the proof, we made some small sacrifices by further estimating some terms in order to obtain a more readable result — the proof thus yields stronger (and asymptotically accurate) estimates. By way of a numerical example, if $n = 100$, $k_1 = 20$ and $k_2 = 40$, the logarithms of the three quantities appearing in Theorem 4 are

$$65.4169 < 65.7549 < 68.0469,$$

where the closeness of the first inequality is in line with expectations since the first inequality in Theorem 2 is quite a bit tighter than the one used for the upper estimate of the sum.

Proof. Applying Lemma 3 with k in the range $2 \leqslant k_1 \leqslant k_2 \leqslant n/2$ and $k_1 - 1 = \alpha_1 n$, $k_2 = \alpha_2 n$, we obtain the following lower estimate for a binomial sum

$$\sum_{k=k_1}^{k_2} \binom{n}{k} > 2^{n+\frac{1}{2}} \exp\left(\frac{1}{1+12n} - \frac{n}{12(n-1)}\right) \sqrt{\frac{n}{\pi}} \int_{\alpha_1}^{\alpha_2} \left(1 - 2\left(\alpha - \frac{1}{2}\right)^2\right)^n d\alpha$$

where we also used the trivial estimates $\alpha(1-\alpha) \leqslant \frac{1}{4}$ and $\alpha \geqslant \frac{1}{n}$ to obtain a cleaner expression. As a final clean-up, the substitution $\alpha - \frac{1}{2} = \frac{1}{\sqrt{2}} \sin t$ turns the latter estimate into

$$\sum_{k=k_1}^{k_2} \binom{n}{k} > 2^n \exp\left(\frac{1}{1+12n} - \frac{n}{12(n-1)}\right) \sqrt{\frac{n}{\pi}} \int_{\beta_1}^{\beta_2} \cos^{2n} t \, dt,$$

where $\beta_1 = \arcsin\left(\sqrt{2}\left(\frac{k_1-1}{n} - \frac{1}{2}\right)\right)$ and $\beta_2 = \arcsin\left(\sqrt{2}\left(\frac{k_2}{n} - \frac{1}{2}\right)\right)$, and $n \geqslant 1$ (note that $\beta_j \in (-\pi/4, 0)$).

Concerning the upper estimate, we again apply Lemma 3 and we first obtain

$$\sum_{k=k_1}^{k_2} \binom{n}{k} < \exp\left(\frac{1}{12n} - \frac{2}{1+6n}\right) \sqrt{\frac{n}{2\pi}} \int_{\alpha_1}^{\alpha_2} \frac{1}{\sqrt{\alpha(1-\alpha)}} \left(1 + 2\sqrt{\alpha(1-\alpha)}\right)^n d\alpha$$

where $1 \leqslant k_1 \leqslant k_2 < n/2$ and $k_1 = \alpha_1 n$, $k_2 + 1 = \alpha_2 n$. Finally, the substitution $\alpha = \frac{1}{2}(1 + \sin t)$ yields $\alpha(1-\alpha) = \frac{1}{4}\cos^2 t$ and

$$\sum_{k=k_1}^{k_2} \binom{n}{k} < 2^n \exp\left(\frac{1}{12n} - \frac{2}{1+6n}\right) \sqrt{\frac{2n}{\pi}} \int_{\gamma_1}^{\gamma_2} \cos\left(\frac{t}{2}\right)^{2n} dt$$

where $\gamma_1 = \arcsin\left(2\left(\frac{k_1}{n} - \frac{1}{2}\right)\right)$ and $\gamma_2 = \arcsin\left(2\left(\frac{k_2+1}{n} - \frac{1}{2}\right)\right)$, and $n \geqslant 1$ (note that $\gamma_j \in (-\pi/2, 0)$). $\square$

3. Appendix: Proof of Theorem 1

Proof of Theorem 1. We begin by proving the first inequality in (2). Taking logarithms left and right and considering the difference, our attention shifts to the function

$$g(x) := -\ln\left(1+4x-4x^2\right) - x\ln x - (1-x)\ln(1-x).$$

We want to show that $g(x) \geqslant 0$ with identity if and only if $x \in \{0, \frac{1}{2}, 1\}$. A calculation gives

$$g'(x) = \frac{4(1-2x)}{1+4x-4x^2} + \ln(1-x) - \ln x,$$

$$g''(x) = -\frac{(1-2x)^2(1-12x+12x^2)}{x(1-x)(1+4x-4x^2)^2},$$

and thus $g(x)$ is convex on $[\frac{3-\sqrt{6}}{6}, \frac{3+\sqrt{6}}{6}]$ and concave elsewhere in $[0,1]$. Since $g(0) = 0$,

$$g\left(\frac{3-\sqrt{6}}{6}\right) > 0.018 > 0,$$

and $g(x)$ is concave on $(0, (3-\sqrt{6})/6)$, clearly $g(x) > 0$ for $x \in (0, \frac{3-\sqrt{6}}{6}]$. Further, $g'\left(\frac{1}{2}\right) = 0$ and, since $g\left(\frac{1}{2}\right) = 0$, convexity of $g(x)$ on $[\frac{3-\sqrt{6}}{6}, \frac{1}{2})$ implies that $g(x) > 0$ on this latter interval. Therefore, since $g(1-x) = g(x)$ we have that $g(x) \geqslant 0$ on $[0,1]$ with identity if and only if $x \in \{0, \frac{1}{2}, 1\}$ as stated.

We now move on to the proof of the second inequality in (2), which is going to be a bit tougher to verify. We need to show that the function

$$f(x) := x^{4/3} + (1-x)^{4/3} - \frac{1}{(1+4x-4x^2)^{1/3}} \tag{3}$$

is non-negative on $[0,1]$ with identity if and only if $x \in \{0, \frac{1}{2}, 1\}$. Using the binomial series gives the representation

$$f(x) = 2^{-1/3} \sum_{k=1}^{\infty} \left[\binom{4/3}{2k} 2^k - (-1)^k \binom{1/3}{k}\right] 2^k \left(x - \frac{1}{2}\right)^{2k}$$

(which converges absolutely for $x \in [0,1]$). A calculation shows that the expression

$$m(k) := \binom{4/3}{2k} 2^k - (-1)^k \binom{1/3}{k} \tag{4}$$

is positive for $k = 1$, negative for $2 \leqslant k \leqslant 7$ and positive for $k = 8$. In fact, we now claim it is positive for all $k \geqslant 8$. We proceed by induction. First, when $k = 8$ (4) is a difference of two positive terms and yields

$$m(8) = \frac{4522537472}{31381059609} - \frac{5434}{59049} > 0.05 > 0.$$

Assume then that $m(k) > 0$. Since we have

$$\binom{4/3}{2(k+1)}2^{k+1} = \left(\binom{4/3}{2k}2^k\right)\cdot\frac{(2k-\frac{4}{3})(2k-\frac{1}{3})}{(2k+1)(k+1)},$$

$$(-1)^{k+1}\binom{1/3}{k+1} = \left((-1)^k\binom{1/3}{k}\right)\cdot\frac{k+\frac{1}{3}}{k+1}$$

and (for $k > 2$, hence in all our relevant cases)

$$\frac{(2k-\frac{4}{3})(2k-\frac{1}{3})}{(2k+1)(k+1)} > \frac{k+\frac{1}{3}}{k+1},$$

our claim is proved. Now, consider the Taylor polynomial of degree 14 for $f(x)$ at $x=\frac{1}{2}$:

$$\tilde{f}(x) := 2^{-1/3}\sum_{k=1}^{7}\left[\binom{4/3}{2k}2^k - (-1)^k\binom{1/3}{k}\right]2^k\left(x-\frac{1}{2}\right)^{2k}.$$

By the claim we just proved, we clearly have $f(x) \geqslant \tilde{f}(x)$ for $x \in [0,1]$ with identity only if $x = \frac{1}{2}$. Now, define the polynomial of degree six

$$F(z) := 2^{-1/3}\sum_{k=1}^{7}\left[\binom{4/3}{2k}2^k - (-1)^k\binom{1/3}{k}\right]2^k z^{k-1}$$

and note that we have

$$\tilde{f}(x) = F\left(\left(x-\frac{1}{2}\right)^2\right)\cdot\left(x-\frac{1}{2}\right)^2.$$

Since $F(0) > 0$ and the coefficients of z^j for $j > 0$ are negative, clearly $F(z)$ is strictly decreasing for $z > 0$ and thus $F(z)$ has exactly one real root. Now, since a numerical calculation shows that $F(0.24) > 0.002 > 0$, reverting to the variable x it follows that $\tilde{f}(x) \geqslant 0$ for all $x \in (\frac{1}{2}-\sqrt{0.24}, \frac{1}{2}+\sqrt{0.24})$, with identity within this interval if and only if $x = \frac{1}{2}$ and, consequently, the same is true for $f(x)$.

At this point let us go back to function $f(x)$ itself (see (3)). We have $f(0) = 0$ and a calculation shows that $f'(0) = 0$. Further, for $x > 0$,

$$f''(x) = \frac{4}{9}\left(\frac{1}{x^{2/3}} + \frac{1}{(1-x)^{2/3}} - \frac{(4-8x)^2}{(1+4x-4x^2)^{7/3}} - \frac{6}{(1+4x-4x^2)^{4/3}}\right)$$

meaning that $f(x)$ is strictly increasing and positive for x small enough (since $\lim_{x\downarrow 0} f''(x) = \infty$ and thus $f(x)$ is initially strictly convex). A numerical calculation now show that

$$\left.\frac{1}{x^{2/3}} + \frac{1}{(1-x)^{2/3}}\right|_{x=\frac{1}{2}+\sqrt{0.24}} > 22.4,$$

while

$$\frac{(4-8x)^2}{(1+4x-4x^2)^{7/3}} - \frac{6}{(1+4x-4x^2)^{4/3}} \leqslant 22 \tag{5}$$

throughout $x \in [0,1]$: indeed the latter observation easily follows from the fact that the function

$$h(w) := \frac{64\left(\frac{1}{4}-w\right)}{(1+4w)^{7/3}} + \frac{6}{(1+4w)^{4/3}}$$

has negative derivative for $w \in [0, \frac{1}{4}]$ and thus the function on the left hand side of (5) is maximal at $x = 0$ (and $x = 1$). Consequently, $f(x)$ is strictly convex through the interval $[0, \frac{1}{2} + \sqrt{0.24}]$, which together with the first part of the proof means that $f(x) > 0$ for all $x \in (0,1) \setminus \{\frac{1}{2}\}$, as we had to prove.

Finally, we need to make sure that the choice of exponents $q = 0$ and $p = \frac{1}{3}$ in Theorem 1 is optimal, that is, that the inequality (2) is no longer true if $q > 0$ or $0 < p < \frac{1}{3}$. To see this, simply observe that the derivative of

$$f_p(x) := x^{p+1} + (1-x)^{p+1} - \frac{1}{(1+4x-4x^2)^p}$$

at $x = 0$ is equal to $3p-1$, and thus since $f_p(0) = 0$ the function $f_p(x)$ is negative for $x > 0$ small enough when $p < \frac{1}{3}$. So could it be that, perhaps, for some $0 < p < \frac{1}{3}$ we have $f_p(x) \leqslant 0$ throughout $[0,1]$? The answer is no, because $f_p(\frac{1}{2}) = f_p'(\frac{1}{2}) = 0$ and yet

$$f_p''\left(\frac{1}{2}\right) = p^2 2^{2-p} > 0,$$

that is, $f_p(x)$ will always be strictly convex and have zero derivative at $x = \frac{1}{2}$, meaning that $f_p(x) > 0$ for x close enough to $\frac{1}{2}$. This concludes the proof of Theorem 2. □

REFERENCES

[1] D. S. MITRINOVIĆ, *Analytic Inequalities*, Springer-Verlag, 1970.

[2] D. J. C. MACKAY, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[3] H. ROBBINS, *A remark on Stirling's formula*, Amer. Math. Monthly **62** (1955), 26–29.

[4] P. STĂNICĂ, *Good lower and upper bounds on binomial coefficients*, Journal of Inequalities in Pure and Appl. Math. **2** (2001), Issue 3, Article 30 (`http://www.emis.de/journals/JIPAM/`).

(Received February 4, 2012)

Vania Mascioni
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306-0490, USA
e-mail: vmascioni@bsu.edu